Correlation and Causality in Genetic Networks

Mary J. Dunlop[1*], R. Sidney Cox[2], Richard M. Murray[3], and Michael B. Elowitz[4]

***Short Abstract* — We present a method for determining the structure of a genetic network given temporal measurements of gene expression. Correlations in time series data are used to determine which genes influence each other and their causal relationships. Natural stochastic noise is shown to aid in the process of network identification by perturbing the expression of genes; the speed and direction at which the noisy signal propagates shows how the network is connected. Mathematical models of simple genetic networks demonstrate that network inference based on correlation data is possible. A library of synthetic genetic circuits is being developed to test these predictions experimentally.**

I. Extended Abstract

A focus of recent work in molecular biology is understanding the function of genetic networks based on their structure [1]. However, finding the structure of a genetic network is a challenging problem and predictive models for network function rely on its structure being well characterized.

Methods for identifying the connectivity of regulatory networks have improved as genetic assays have advanced. Large-scale network identification has focused on steady state measurements and statistical inference algorithms [2-5]. Alternative approaches for smaller-scale networks have been suggested using temporal data [6-7].

We perform network identification by analyzing the cross correlation between time series measurements of two genes. If a signal takes time to propagate from one gene to the next this appears as a lag in the cross correlation function; the type of interaction (activation or repression) is reflected in the sign. The cross correlation measurement reveals the causality, or direction of interaction, between two genes. A similar approach was applied previously in metabolic networks [8], where the input to the chemical system was accurately controlled. Network identification algorithms benefit from perturbing the system and measuring its response [5, 6]. Examples of perturbations include changing initial conditions, varying the environment in which the network is measured, or altering genetic pathways through gene knock outs. We take a different approach and rely on noise in gene expression as a perturbing force. Traditionally a hindrance, we demonstrate numerically that stochastic gene expression is a sufficient perturbation to allow for discrimination between network topologies. In parallel, we are constructing synthetic genetic networks to test these network inference methods experimentally.

Networks are modeled using sets of ordinary differential equations with Hill functions representing transcriptional control [1]. Model parameters from previous experiments are used [9]; noise is modeled with an Ornstein-Uhlenbeck process [10] and both intrinsic and extrinsic noise sources are included. Model results show that networks with different connectivities can be distinguished using noise alone to perturb the system.

We are testing these methods experimentally with a library of synthetic genetic networks with known circuit topologies. Quantitative measurements are taken in single cells and traces of fluorescence versus time in each cell are recorded.

Both numerical and preliminary experimental results will be presented.

Acknowledgements: Research supported in part by the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office.

[1, 3] Division of Engineering and Applied Science. California Institute of Technology, Pasadena, CA. 91125, USA.

[2,4] Division of Biology, California Institute of Technology, Pasadena, CA. 91125. USA.

[*] Email: mjdunlop@caltech.edu.

References

[1] U. Alon. An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman and Hall, 2007.

[2] Friedman, N.; Linial, M.; Nachman, I. & Pe'er, D. "Using Bayesian networks to analyze expression data." Journal Of Computational Biology, 2000, 7, 601-620.

[3] Gardner, T. S.; di Bernardo, D.; Lorenz, D. & Collins, J. J. "Inferring genetic networks and identifying compound mode of action via expression profiling." Science, 2003, 301, 102-105.

[4] Faith, J. J.; Hayete, B.; Thaden, J. T.; Mogno, I.; Wierzbowski, J.; Cottarel, G.; Kasif, G.; Collins, J. J. & Gardner, T. S. "Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles." PLoS Biology, 2007, 5, 54-66.

[5] Geier, F.; Timmer, J. & Fleck, C. "Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge." BMC Systems Biology, 2007.

[6] Sontag, E.; Kiyatkin, A. & Kholodenko, B. N. "Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data." Bioinformatics, 2004, 20, 1877-1886.

[7] de Hoon, M.; Imoto, S. & Miyano, S. "Inferring gene regulatory networks from time-ordered gene expression data using differential equations." Discovery Science, Proceedings, 2002, 2534, 267-274.

[8] Arkin, A.; Shen, P. D. & Ross, J. "A test case of correlation metric construction of a reaction pathway from measurements." Science, 1997, 277, 1275-1279.

[9] Rosenfeld, N.; Young, J. W.; Alon, U.; Swain, P. S. & Elowitz, M. B. "Gene regulation at the single-cell level." Science, 2005, 307, 1962-1965.

[10] Uhlenbeck, G. E. & Ornstein, L. S. "On the theory of the Brownian motion." Physical Review, 1930, 36, 0823-0841.